

ORIGINAL ARTICLE

Phylogenetics and movement of Phycodnaviruses among aquatic environments

Manuela V Gimenes¹, Paolo M de A Zanotto¹, Curtis A Suttle², Hillândia B da Cunha³ and Dolores U Mehner¹

¹Department of Microbiology, University of São Paulo, São Paulo, Brazil; ²Department of Earth and Ocean Sciences, University of British Columbia, Vancouver, Canada and ³National Institute of Amazon Research (INPA), Manaus, Brazil

Phycodnaviruses have a significant role in modulating the dynamics of phytoplankton, thereby influencing community structure and succession, nutrient cycles and potentially atmospheric composition because phytoplankton fix about half the carbon dioxide (CO_2) on the planet, and some algae release dimethylsulphoniopropionate when lysed by viruses. Despite their ecological importance and widespread distribution, relatively little is known about the evolutionary history, phylogenetic relationships and phylodynamics of the Phycodnaviruses from freshwater environments. Herein we provide novel data on Phycodnaviruses from the largest river system on earth—the Amazon Basin—that were compared with samples from different aquatic systems from several places around the world. Based on phylogenetic inference using DNA polymerase (*pol*) sequences we show the presence of distinct populations of *Phycodnaviridae*. Preliminary coarse-grained phylodynamics and phylogeographic inferences revealed a complex dynamics characterized by long-term fluctuations in viral population sizes, with a remarkable worldwide reduction of the effective population around 400 thousand years before the present (KYBP), followed by a recovery near to the present time. Moreover, we present evidence for significant viral gene flow between freshwater environments, but crucially almost none between freshwater and marine environments.

The ISME Journal (2012) 6, 237–247; doi:10.1038/ismej.2011.93; published online 28 July 2011

Subject Category: microbial population and community ecology

Keywords: Amazon; *Phycodnaviridae*; phylodynamics; phylogeny

Introduction

The family *Phycodnaviridae* comprises six genera that include several large and diverse icosahedral double-stranded DNA viruses, with genomes ranging from 160 kbp to 560 kbp (Van Etten *et al.*, 2002; Dunigan *et al.*, 2006; Wilson *et al.*, 2008), which infect a wide spectrum of eukaryotic algae. Phycodnaviruses have a significant role in modulating the dynamics of phytoplankton, thereby influencing the community structure and succession (Castberg *et al.*, 2001; Brussaard, 2004), nutrient cycles (Gobler *et al.*, 1997; Wilhelm and Suttle, 1999; Rusch *et al.*, 2007) and potentially atmospheric composition. This is because phytoplankton fix about half of the CO_2 on the planet (Sabine *et al.*, 2004; Denman *et al.*, 2007; Houghton, 2007), and some algae release dimethylsulphoniopropionate when lysed by viruses (Liss *et al.*, 1997; Hill *et al.*, 1998; Malin *et al.*, 1998). Despite their ecological importance and

widespread distribution (Cottrell and Suttle 1991; Clasen and Suttle, 2009), relatively little is known about the evolutionary history and phylogenetic relationships among the Phycodnaviruses (Chen and Suttle, 1996; Short and Suttle, 2002; Wilson *et al.*, 2006; Larsen *et al.*, 2008) particularly as they pertain to freshwater environments (Short and Short, 2008; Clasen and Suttle, 2009). Moreover, it has been postulated that, based on molecular phylogenies, viruses and other microorganisms have experienced a limited number of marine-freshwater transitions during their evolution (Logares *et al.*, 2009).

The *Phycodnaviridae* family was recognized based on a group of viruses that infect endosymbiotic Chlorella-like algae (Van Etten and Ghabrial, 1991), and is now known to be part of the large nuclear and cytoplasmic DNA virus group. These viruses encode several conserved proteins performing most key life-cycle processes. The choice for using the B-family DNA polymerase as a conserved marker for the evolutionary history of nuclear and cytoplasmic DNA virus group, and its use as a target in related studies, lay on the fact that this protein contains a polymerase domain that is highly conserved at the amino acid level, and less conserved at the nucleotide level, allowing for deeper phylogenetic inference (Zhang and Suttle, 1994).

Correspondence: MV Gimenes, ICB-II, Universidade de São Paulo, Av Prof Lineu Prestes, 1374, Cidade Universitária, São Paulo, SP 05508-900, Brazil.

E-mail: manuela_vg@hotmail.com

Received 18 April 2011; revised 6 June 2011; accepted 6 June 2011; published online 28 July 2011

The Amazon is one of the most ancient freshwater environments, raising the issue if it could function as a reservoir from which freshwater viruses could disperse to other environments. Despite the magnitude of the Amazonian river system little is known in general about the microbes that inhabit these waters (Benner *et al.*, 1995; Hungria *et al.*, 2005; Rejas *et al.*, 2005; Hewson *et al.*, 2006). Particularly, there is a lack of studies on Phycodnaviruses in tropical freshwater environments in general and in the Amazonian river system in particular. This study constitutes the first report on the composition of Phycodnaviruses of tropical waters in South America from the Amazon Basin, which is the source of 20% of the free-flowing freshwater on Earth, as well as being one of the world's most important ecosystems in terms of biodiversity.

Materials and methods

Study sites

The two main rivers of the Amazon Basin are the Solimões and the Negro, which, after meeting near the city of Manaus, form the great Amazon River. Despite being geographically close, both rivers are physically, chemically and biologically distinct. The Solimões River is considered as a white-water river, having brownish turbid waters, rich in particulate suspension material and pH ranging from 6.2 to 7.2 (Sioli, 1975a, 1975b). The Negro River, on the contrary, is considered a black-water one. The dark brown color of the water results from the high concentration of humic substances in it, originating from the decomposition of high quantities of organic matter present in the river. Its acidic waters (pH from 3.8 to 4.9) have low concentrations of dissolved salts, reflecting the chemical poverty of the soils the waters pass over (Sioli, 1975a, 1975b). Both the Negro River, and its tributary, the Cuieiras River, share the same black-water-river characteristics.

However, the Cuieiras River flows over an area with almost no human impact.

Sample collection and viral concentration

Freshwater samples were collected from the Solimões ($n=3$), Negro ($n=4$) and Cuieiras ($n=1$) Rivers (Table 1), during the flood of 2007 (July), using a submersible pump. The water was transported to the laboratory in 20-l polypropylene carboys, protected from light and heat and processed within 2–3 h. A volume of 10–100 l of water were pressure filtered through 142 mm diameter cellulose fiber paper filter (Millipore, Barueri, Brazil), 142 mm diameter 1.2 µm pore size glass fiber (Millipore) and 142 mm diameter 0.45 µm pore size PVDF (Millipore) membrane filters connected in series. The remaining particulate material was concentrated ~1000-fold by ultrafiltration through a 30-kDa-cutoff tangential flow cartridge (Millipore) (Suttle *et al.*, 1991) followed by ultracentrifugation at 100 000 g for 2 h in a swinging bucket rotor (Mehnert and Stewien, 1993). Pellets were resuspended in supernatant and stored at 4 °C in the dark until the DNA was extracted.

Phycodnaviruses DNA amplification

After treatment with VerteXF (DuPont, Barueri, Brazil) to remove lipids and proteins (Queiroz *et al.*, 2001), DNA from each concentrate was extracted using the PowerSoil DNA kit (MoBio, Cotia, Brazil) and stored at –20 °C until use. Phycodnavirus DNA polymerase (*pol*) gene fragments were amplified in two PCR rounds with the degenerate primers, AVS1/AVS2 and AVS1/POL (Chen and Suttle, 1995), using a Hybaid PCR Express thermocycler (Ashford, UK). In the first round 2 µl of template DNA were added to a mixture of 1 U of Platinum *Taq* DNA polymerase (Invitrogen, Burlington, Canada), 0.2 mM of each deoxyribonucleotide triphosphate, 10 pmol of AVS1, 60 pmol of AVS2, 1.5 mM MgCl₂,

Table 1 Details of samples used in this study

Sample	Date 2007	Location	Coordinates	Depth (m)	Current (km h ⁻¹)	Water temperature (°C)	pH	DO (mg l ⁻¹)	Volume (l)	
									Initial	Final
SOLa	07/18	Solimões River—Ilha da Machantaria	03°16'S 60°02'W	1.5	5.7	27.8	6.5	2.36	10	0.01
SOLb	07/23			1.5	5.6	27.9	6.2	2.64	20	0.02
SOLc	07/26			1.5	5.9	27.9	6.0	2.84	20	0.02
CUIa	07/24	Cuieiras River—at the source	02°46'S 60°27'W	1.5	0	27.8	4.0	3.2	80	0.08
NEGa	07/17	Negro River—Tatu	03°03'S 60°18'W	1.5	3.5	28.7	4.7	3.12	100	0.10
NEGb	07/20			1.5	4.6	28.3	4.5	3.12	100	0.10
NEGc	07/25			1.5	3	28.6	4.2	3.23	100	0.10
NEGd	07/19	Negro River—Educandos	03°03'S 60°01'W	1.5	0	28.7	4.8	2.6	40	0.04

Abbreviation: DO, dissolved oxygen.

a, b, c, d indicate the different samples from the same river.

manufacturer-provided PCR reaction buffer and water, to a final volume of 50 µl. In the negative and positive control reactions, template DNA was substituted for water and *Micromonas pusilla* virus SP-1 (MpV-SP1) DNA, respectively. PCR conditions were 90-s initial denaturation at 95 °C, followed by 40 cycles of 45-s denaturation at 95 °C, 45-s annealing at 45 °C and 45-s extension at 72 °C, and plus a final 10-min extension.

A pool of four PCR reactions from each sample (the total volume of 200 µl) was electrophoresed with 6 × loading buffer on a 1.5% agarose gel, at 90 V for 60 min, in 0.5 × Tris-Borate-EDTA buffer. The ethidium bromide-stained gel was visualized with ultraviolet for <5 s using an AlphaImager 3400 (AlphaInnotech, Toronto, Canada). The ~700 bp bands were excised, and then purified using a MinElute PCR Purification kit (Qiagen, Toronto, Canada) before elution in 50 µl of water.

A nested second-round PCR was performed, to confirm that the correct target was amplified. This was done by adding 2 µl of purified DNA to a mixture of 0.625 U of Platinum *Taq* DNA polymerase, 0.16 mM of each deoxyribonucleotide triphosphate, 30 pmol of both AVS1 and POL primers, 1.5 mM MgCl₂ and manufacturer-provided PCR reaction buffer and water to a final volume of 25 µl. PCR conditions were a 90-s initial denaturation, followed by 30 cycles of 30-s denaturation at 95 °C, 45-s annealing at 50 °C and 60-s extension at 72 °C, plus a final 10-min extension at 72 °C. Negative and positive controls were the same as in the first round amplification. A subsample of 100 µl from the ~500 bp PCR products of each sample (a pool of four reactions) was electrophoresed, purified, eluted and stored as indicated above.

Cloning and sequencing

Purified 500 bp *pol* fragments were cloned into pCR4-TOPO vectors (Invitrogen). *Escherichia coli* strain TOP-10 chemocompetent cells were transformed, and plated on Luria-Bertani/agar medium selective for ampicillin (100 µg ml⁻¹), according to the manufacturer's protocol. Bacterial colonies containing *pol* inserts were randomly picked and inoculated in 80 µl of Luria-Bertani with ampicillin (100 µg ml⁻¹) in 96-well plates. After a 12-h incubation period at 37 °C, 40 µl of sterile 60% glycerol were added to each well. Plates were incubated for an additional hour at 37 °C and stored at -80 °C until sent to the Genome Québec Innovation Centre at McGill University for sequencing on a 3730xl DNA Analyzer (Applied Biosystems, Carlsbad, CA, USA).

Phylogenetics and phylogeographic analysis

Sequence alignment. A data set of B-family (α -like) Phycodnavirus DNA *pol* sequences from cultures and environmental samples, including the ones from this study, was aligned by introducing gaps

respecting the proper coding frame with Clustal X (Thompson *et al.*, 1997) and Codon Align (<http://www.sinauer.com/hall/2e/>).

Phylogenetic analysis. At first, a global maximum likelihood (ML) tree of *pol* sequences was constructed from an alignment of 244 inferred amino-acids from 638 taxa, including environmental sequences (Supplementary information 1) and representatives from all *Phycodnaviridae* genera: *Paramecium bursaria Chlorella virus*—PBCV (*Chlorovirus*), *Chrysochromulina brevifilum virus*—CbV and *Phaeocystis globosa virus*—PgV (*Prymnesiovirus*), *Ectocarpus siliculosus virus*—EsV (*Phaeovirus*), *Emiliania huxleyi virus*—EhV (*Coccolithovirus*), *Heterosigma akashiwo virus*—HaV1 (*Raphidovirus*), *M. pusilla virus*—MpV (*Prasinovirus*) and *Ostreococcus virus*—OsV and *Bathycoccus virus*—BpV (unclassified) (Supplementary information 2). To characterize and better define deep associations among Phycodnaviruses, a diverse set of taxa were used as outgroups (HaV1, EhV, EsV, PgV, CbV and PBCV) during protein-based phylogenetic inferences.

Protein trees were recovered with PhyML v2.4.4 (Guindon and Gascuel, 2003) using the JTT model of aminoacid substitution, with the following parameters optimized from the data: (i) proportion of invariable sites, (ii) tree topology and (iii) shape parameter (α) of the γ -distributed substitution rates. Additionally, in a second analysis using DNA sequences, the outgroups were removed to better estimate the associations among closely related groups by reducing saturation at degenerate sites introduced by highly divergent sequences. DNA-based phylogenies for Phycodnaviruses alone were obtained from the analysis of complete data sets using BEAST v.1.4.8 (Drummond and Rambaut, 2007; see details below). Maximum clade credibility trees summarizing the posterior probability of nodes on the phylogeny were obtained from trees sampled at stationary during the Markov Chain Monte Carlo (MCMC) runs.

Phyldynamic analysis. Based on the coalescent theory of Kingman (1982), intra-species gene genealogies have been extensively used to infer various demographic parameters for a diverse set of organisms (Page and Holmes, 1998), allowing the time and mode of evolution of exogenous viruses (Zanotto *et al.* 1996, Grenfell *et al.*, 2004, Pybus and Rambaut, 2009), endogenous viruses (Romano *et al.*, 2007) and even complex metazoa (Campos *et al.*, 2010) to be understood. To infer gene genealogies closely related *pol* sequences were subdivided into demes. We considered demes as discrete subpopulation of viruses with definable genetic relationships. As not much is known about any Phycodnavirus population structure and nor how comprehensive our samples would be of any of the viral 'species' we sampled from, we partitioned our sequences into demes (that is, subpopulations)

following three criteria: (i) they had to belong to a monophyletic group (that is, they had to share a common relatively recent ancestral node), (ii) have 17 or more samples, which implies a probability of 0.0074 of adding any new sequence that would alter the time to the most recent common ancestor (that is, adding 136 new sequences per deme, Felsenstein, 1993) and (iii) not show genetic divergence within the monophyletic group >2%. The average pairwise genetic distance among taxa from apical monophyletic groups found in the global tree was calculated from the ML estimates of the number of substitutions per site obtained with the DNADIST program version 3.6, available in the PHYLIP package (Felsenstein, 1993).

Demes were analyzed using a Bayesian MCMC approach implemented in BEAST. Bayesian priors for the mutation rate per site per year for Phycodnaviruses were approximated using the formula: $\log(\text{rate}) = 0.95 \log (\text{genome size}) + 2.67$ ($r^2 = 0.97$), obtained by interpolating the genome size ranges for Phycodnaviruses (160–560 kbp) into the per-site mutation rate versus genome size curve for organisms ranging from viroids to higher eukaryotes (Gago *et al.*, 2009, data available from authors upon request). A relaxed (uncorrelated lognormal) molecular clock, with a normal prior distribution of mean = 1.4E-8 s/s/y (substitutions per site per year) and s.d. = 2.34E-7 s/s/y was used, as dates of sampling would be uninformative for this time scale. This rate obtained by interpolation agrees with rates of substitution per site per year for double-stranded DNA viruses (Duffy *et al.*, 2008), such as Herpes virus type 1, with a rate of 1.8E-8 s/s/y (Drake and Hwang, 2005) and a genome size within the same order of magnitude of the Phycodnavirus. Therefore, we believe that the prior distribution we used would include realistic values of the rate of change per site per year. The behavior of the Phycodnavirus population size through time was inferred using virus sequences and a prior distribution for the mutation rate. This is because, based on phylogenies of Phycodnaviruses sampled during the MCMC, BEAST infers the population size parameter based on the number of cladogenetic events (that is, nodes or bifurcation events along the tree from the tips back to the root) to estimate the joint parameter ($N_e \times g$) that is, effective population size (N_e) times the generation time in years (g). Nevertheless, from our data or prior knowledge of our system, we do not have ways to disentangle the actual values of effective population size from generation time, but merely report their joint behavior in time, as shown in Bayesian skyline plots.

To be as conservative as possible an unconstrained Bayesian skyline coalescent prior was used under the best-fit model of nucleotide substitution (HKY + Γ + I, Shapiro *et al.*, 2006). To improve the MCMC search operator values were set to a function of the number of taxa in each group. After optimizing the values of the MCMC operators during

preliminary runs, up to 10 additional MCMC runs individually, consisting of 20 million generations, were done to obtain convergence of parameter estimates, indicated by values of effective sampling sizes above 200. In all cases the convergence of parameters during the MCMC was inspected with Tracer v.1.4 (Drummond and Rambaut, 2007), with uncertainties depicted as 95% high probability density (HPD) intervals. To provide an independent assessment of the robustness of our analysis of evolutionary dynamics, the time to the most recent common ancestor of each data set was also estimated using the Path-O-Gen software (available at <http://tree.bio.ed.ac.uk/software/pathogen/>), which uses a linear regression of root-to-tip genetic distances against sampling time, based on the ML phylogenies described above.

Phycodnaviruses movement among aquatic environments. The *BayesTraits* software (Pagel and Meade, 2004) was used to derive posterior distributions and ML estimates of values for rates of change among traits, which were assumed as posterior probability estimates of both the rates of movement (that is, migration) among the three environments (riverine, lacustrine and marine) and the posterior probability of the state at the root of the trees (that is, origin of the viruses). Analyses were performed using 10 000 trees sampled after the MCMC reached stationarity (usually after 20 million states), and by coding each taxa as discrete multistate traits representing their environmental sources ('river' = R, 'lake' = L and 'sea' = S). Several freshwater samples could not be defined as either R or L; therefore the ambiguity character coding RL (that is, either 'river' or 'lake') was used instead (Pagel and Meade, 2004). Missing traits were coded as hyphens. To improve the convergence of estimates, at least 100 attempts in finding the likelihood for each tree with *BayesTrait* were used.

Results

Sequences obtained from the eight water samples collected in this study yielded 65 and 39 Phycodnavirus *pol* sequences from the Solimões and Cuieiras Rivers, respectively, providing the first data on Phycodnaviruses from tropical rivers. These sequence data have been submitted to the GenBank database under accession numbers HQ424349–HQ424430. This data set allowed us to do the first comparison between riverine Phycodnaviruses clones from both hemispheres and from temperate and tropical environments.

Molecular characterization of Amazonian Phycodnaviruses

No Phycodnavirus amplicons were obtained from the Negro River samples. The lack of

Phycodnaviruses in the Negro River could be because of naturally occurring conditions because we did observe phages by electron microscopy but not Phycodnaviruses (data not shown). It is likely that these viruses were either missing at the spot or below the level of detection of our method. Nevertheless, DNA was isolated and sequences from other viruses could be amplified from these samples (data not shown). The physical and chemical conditions of these waters were very similar to those found in the Cuieiras River, from where Phycodnavirus amplicons were obtained. We argue that, possibly, the unique conditions of the river were responsible for the lack of viral detection, assuming that there was no unanticipated methodological problem specific to Phycodnaviruses at those particular sampling localities. All 104 Amazonian sequences were aligned to 550 *pol* sequences from environmental samples, and to sequences from cultured isolates, to create a 638-taxa data set. Phylogenetic analysis with both DNA and inferred amino acid sequences yielded trees that were similar to those inferred in previous studies (Larsen *et al.*, 2008; Short and Short, 2008; Clasen and Suttle, 2009, Figure 1). Our choice of outgroups for the phylogenetic analysis (Figure 1) was justified as these taxa were more divergent than all Phycodnaviruses we sequences or used as references, in line with previous studies

(Short and Suttle, 2003; Larsen *et al.*, 2008; Short and Short, 2008; Clasen and Suttle, 2009). Furthermore, to minimize the impact of saturation at the degenerate coding sites at the DNA level and to properly depict the divergence among the more recently diverged ingroup we excluded the divergent outgroups (HaV1, EhV, EsV, PgV, CbV and PBCV). Nevertheless, our phylogenies revealed highly divergent monophyletic groups showing intense cladogenesis near the tips of the tree branches that are consistent with demographic expansions of closely related populations (Figure 2). These monophyletic groups were identified as demes (shown as yellow sectors in the tree) for the purpose of subsequent demographic studies. The branching pattern on the tree also supports a distant relationship between Phycodnavirus clone demes from freshwater and marine lineages (Figure 2).

Population dynamics of Phycodnaviruses

Population size estimates shown in the Bayesian skyline plot for all Phycodnaviruses (Figure 3) were calculated as the effective population times the generation time ($Ne \times g$). There was a significant reduction in $Ne \times g$ and the 95% HPD reached a nadir between 500 and 300 thousand years before present (KYBP), suggesting a complex

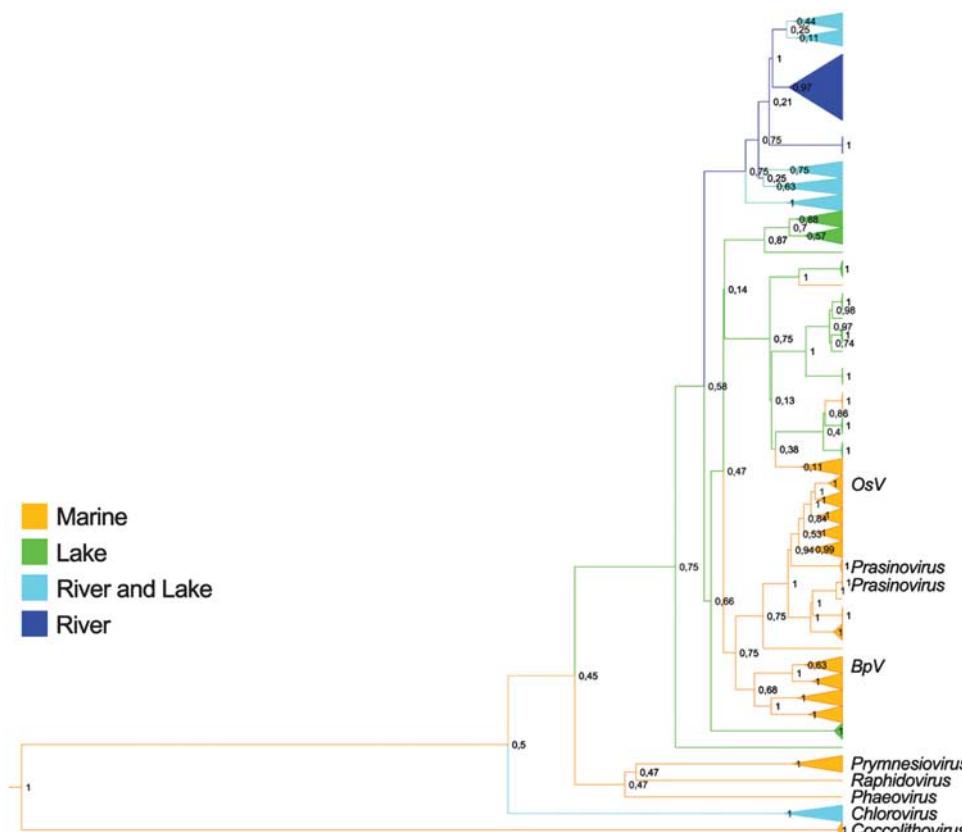


Figure 1 ML tree of Phycodnaviruses based on amino-acid sequences inferred from DNA *pol* sequences. Nodes marked in orange contain marine sequences, nodes marked in light blue contain sequences from lakes and rivers, nodes marked in green contain sequences from lakes and nodes marked in dark blue contain sequences from rivers.

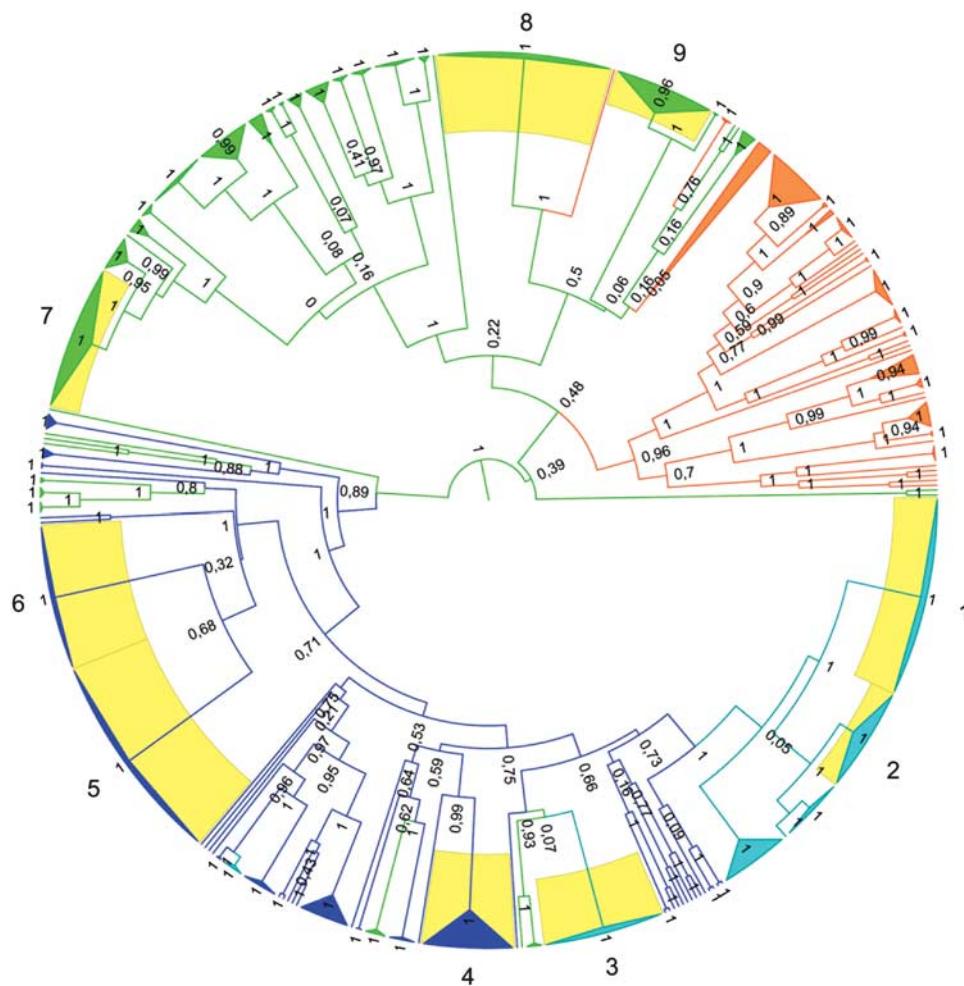


Figure 2 ML tree of Phycodnavirus DNA *pol* inferred amino acid sequences. Nodes are identified by their origin as marine (orange), riverine (dark blue), lacustrine (green) and fresh water (light blue), with sequences from both rivers and lakes. Demes (yellow) are identified by numbers from 1 to 9. Deme 1 and 2—samples from the Chatfield Reservoir (CO, USA) and the South Platte River (CO, USA); Deme 3—samples from the Chatfield Reservoir, the South Platte River and Lake Ontario (ON, Canada); Deme 4—samples from the Solimões River (AM, Brazil); Deme 5—samples from the South Platte River; Deme 6—samples from the Cuieiras River (AM, Brazil); Demes 7 and 9—samples from Lake Ontario and Deme 8—samples from Lake Ontario and Crawford Lake (ON, Canada).

fluctuating dynamics of Phycodnaviruses over a large time scale. The data also provided evidence of a considerable increase in the number of lineages after a bottleneck around 300 KYBP, followed by a plateau around 100 KYBP (Figure 3). This agrees with the signature observed near the present (Figure 4). Moreover, there were no obvious differences in population sizes of demes from the Cuieiras and Solimões Rivers compared with other localities. Deme 2 (from the Chadfield reservoir and the South Platte River) and Deme 4 (from the Solimões River) did not show the population reduction about 300 000 years ago observed elsewhere. Rather, the Solimões River kept a relatively constant size as indicated by a stable $N_e \times g$. Moreover, we noticed that Deme 2 was the only population to increase throughout that period.

Movement of Phycodnaviruses among water environments

To evaluate the apparent restriction of gene flow among Phycodnaviruses, the likelihood method in *BayesTraits* was used to estimate the lineage exchange among aquatic environments. Results indicated significant genetic exchange among rivers and lakes, but exchange rates between marine and fresh waters were very low, ranging from 100 to 10 000-fold less (Figure 5). Significant differences among rates were indicated by a lack of overlap among the 95% HPD estimates, rejecting the hypothesis of random exchange (panmixis) of Phycodnaviruses among marine and fresh waters, and providing evidence for restricted gene flow. Because exchange values between rivers and lakes were the highest, but within the same order of magnitude, we evaluated how the ratio varied

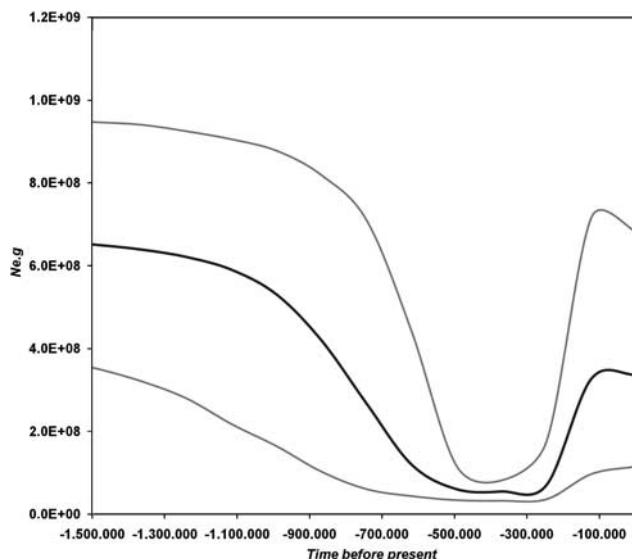


Figure 3 Bayesian skyline plot of all Phycodnaviruses included in this study analyzed together showing the overall dynamics over the last 1.5 million years, during the Quaternary period. The 95% HPD (thin gray lines) are shown as around the mean (thick black line).

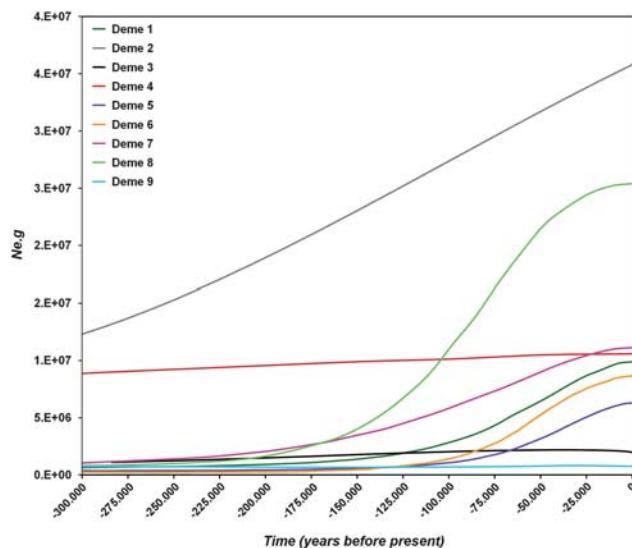


Figure 4 Overlay of Bayesian skyline plot of all Phycodnaviruses' demes summarizing their individual dynamics over the last 300 KYBP. Several demographic signatures show a considerable increase in the number of lineages around 130 KYBP possibly during the warm Eemian interglacial period, reaching a plateau or reduction near the present. Demes 1 and 2—Chatfield Reservoir (USA) and the South Platte River (USA); Deme 3—Chatfield Reservoir (USA), the South Platte River (USA) and Lake Ontario (Canada); Deme 4—Solimões River (Brazil); Deme 5—South Platte River (Canada); Deme 6—Cuieiras River (Brazil); Demes 7 and 9—Lake Ontario (Canada) and Deme 8—Lake Ontario (Canada) and Crawford Lake (Canada).

depending on the tree used for the estimate. Joint estimates of rates obtained with *BayesTraits* from a sample of trees visited by the MCMC method indicated that the ratio was robust to values of gene

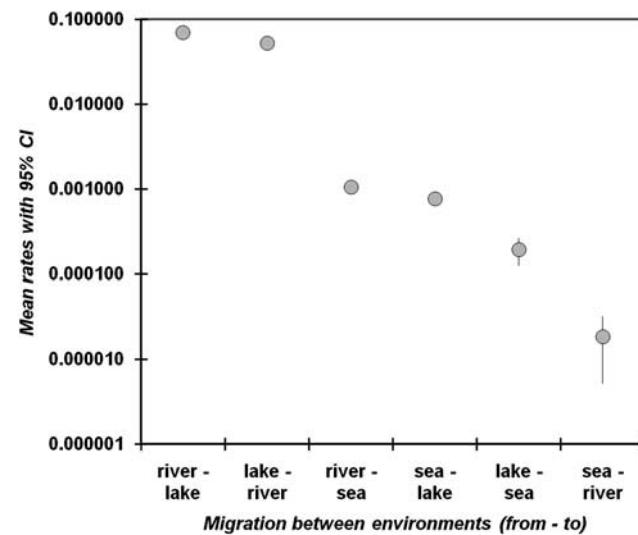


Figure 5 Posterior probability and standard 95% HPD in decreasing order show Phycodnavirus' migration among aquatic environments as the instantaneous rates of change.

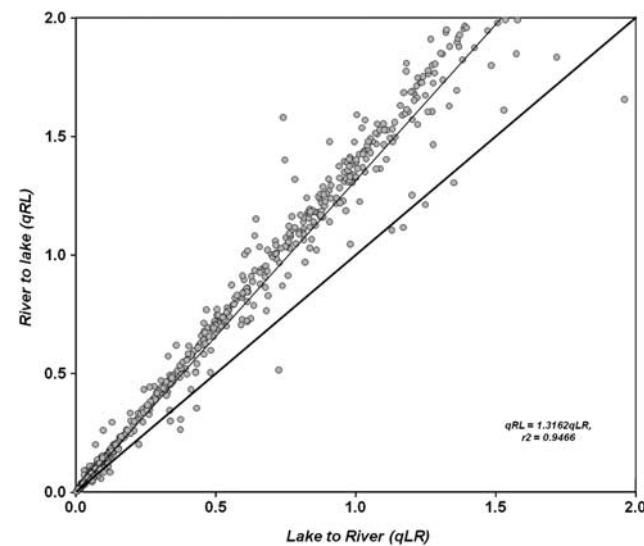


Figure 6 Posterior probability of joint estimates of qRL (indicating rate of movement of Phycodnaviruses from river to lake) and qLR (movement from lake to river) obtained with *BayesTraits* from trees sampled at stationary during MCMC runs. The data indicate that qRL is larger than qLR for most estimates and therefore there appear to be more movement of Phycodnavirus from rivers to lakes.

flux from river to lake, mostly larger than from lake to river (Figure 6). That happened irrespective of the rates co-estimated for each tree that was summarized by the line with a high correlation coefficient ($r^2 = 0.99$) rising above the diagonal. Consequently, this result supports the idea of further asymmetries in Phycodnavirus gene flow among freshwater systems, possibly with greater influx of viruses from 'rivers' into 'lakes' than the other way around. An attempt to infer the root position on MCMC trees

showed higher probabilities for a marine origin (S) of Phycodnaviruses, although large overlapping 95% HPD values for the alternative environments (R, L) indicate that the root habitat is unresolved (data not shown).

Discussion

Evolutionary dynamics of Phycodnaviruses

The phylogenetic trees of DNA *pol* sequences indicated that all samples studied, including those from the Amazon, belong to Group C Phycodnaviruses (Larsen *et al.*, 2008), which are found in a wide range of marine and freshwater environments. The phylogenetics of DNA *pol* sequences was characterized by a complex signature of fluctuating dynamics. It depicted a remarkable reduction of the effective population size ($N_e \times g$) around 400 KYBP, suggestive of a genetic diversity bottleneck with small 95% HPD (Figure 3), indicating a reduction in cladogenetic events before the recent burst of lineages on most demes. We argue that sampling biases alone would not explain this effect. Crucially, increased sampling would not only reduce even further the likelihood of finding a new older root for any particular deme (Felsenstein, 2004), but also magnify the recent growth signal. Moreover, a similar near-present expansion was observed independently in all fresh- and salt-water systems. Therefore, assuming that the pruning of lineages causing the $N_e \times g$ reduction we observed is not due to sampling, it could be explained, to some extent, by a boom and bust process (Ferguson *et al.*, 1999), which causes older lineages to die out within demes (Zanotto *et al.*, 1996). Notably, fluctuations in viral effective sizes may have an important role in controlling the succession dynamics of algal populations, causing abrupt bloom termination (Bratbak *et al.*, 1993; Jacquet *et al.*, 2002). This is part of the well-known boom and bust dynamics of host algal population (that is, 'kill the winner' hypothesis—Thingstad, 2000). In addition, other non-biotic factors could also have had an impact on the overall dynamics we observe.

For example, the gradual reduction of viral effective population size may also be compatible with the known paleontological record. According to our temporal reconstruction of events based on a rate of change of 1.4E-8 s/s/y, the demographic reduction in the number of lineages (Figure 3) appears to have taken place after the Gelasian age, during the Pleistocene, between 1.5 million and 300 thousand years ago. This is a period known for key changes such as the extinction of the calcareous nanofossils *Discoaster surculus* and *D. pentadiatus* and the planktonic foraminifer *Globigerinoides extremus* (Rio *et al.*, 1998; Gradstein *et al.*, 2005). Interestingly, however, after considerable growth in $N_e \times g$ coinciding with the warmer Eemian interglacial period, a decline was observed leading to the present (Figure 3).

Is the environmental movement of Phycodnaviruses dictated by physico-chemistry?

Because we had information related to the place of collection of a diverse set of samples, we did use phylogenetic associations to study the movement of Phycodnaviruses between aquatic environments. Our findings supported the notion that Phycodnaviruses may not be completely mixed (that is, panmictic) with respect to water environments (riverine, lacustrine and marine), and that this gene flow restriction among these environments could have a significant role in their evolution and dispersal. Furthermore, we observed freshwater and marine samples clustering separately, in accordance with previous results obtained from smaller data set (Clasen and Suttle, 2009) in particular and with what has been observed for several other microorganisms in general (Logares *et al.*, 2009). Data on North American Phycodnaviruses did not present good evidence for the deep phylogenetic relationship between marine and freshwater isolates of group I and II as defined by Short and Short (2008). However, our results agree with the finding that freshwater Phycodnaviruses clones are more closely related to some marine isolates (MpV, OsV and BpV) than to freshwater isolate PBCV (Short and Short, 2008). The fact that PBCV radiated deep in the tree, adjacent to highly divergent seawater samples, also hindered a precise placement of the root-tree position, even though the inclusion of sequences from cultivated viruses had placed it at 'sea' in the maximum clade credibility tree.

Our multi-state character change analyses with *BayesTraits* suggested significant restrictions in Phycodnaviruses movement between salt and freshwater, which could possibly explain further restrictions on the movement of these viruses among aquatic environments, or the colonization of specific freshwater systems like the Negro River. On a similar vein, a drastic reduction in freshwater viral counts upon addition of seawater was observed by Cissoko *et al.* (2008), suggestive of salinity restriction for virus movement between freshwater and marine environments. However, the same effect was not observed upon the addition of freshwater to marine samples, indicating that marine viruses can better tolerate reductions in salinity, as do marine bacteria. Similarly, marine phytoplankton are known to cope well with changes in salinity, probably because of their necessity to deal with tidal variation, precipitation and other environmental phenomena (Kirst, 1989), whereas freshwater phytoplankton are usually totally replaced by marine species in a range of 0.5–10‰ salinity (Lionard *et al.*, 2005). Moreover, bacteriophages isolated from lake water and soil were able to replicate in marine microbes (Sano *et al.*, 2004; Logares *et al.*, 2009), which, given that bacteria tolerate greater salinity variations than algae, suggests that the restriction of Phycodnaviruses movement between fresh and salt water environments

may be because of distinctions in host biology to some considerable extent. Nevertheless, freshwater virio, bacterio and phytoplankton seem to be more susceptible to osmolarity stress than the marine phytoplankton. These findings agree with the phylogenetic tree root position in the 'sea', having the highest posterior probability. It is important to have in mind that our main findings and conclusions are based on a limited number of samples and the transition environments between marine and freshwaters. Moreover, as a caveat we should consider that PCR amplification with the highly degenerate primers we used, could generate biased sampling because there is no assurance that they can detect all existing Phycodnaviruses (Chen and Suttle, 1995). Nevertheless, it is worth considering three aspects: (i) this set of primers did amplify samples that nest the Prasinoviruses (for example, our samples had more divergence among themselves than the divergence seen among Prasinoviruses, Figure 1), (ii) the primers we used in this study (AVS1 and POL) successfully amplify the distantly related Phycodnaviruses that infect *E. huxleyi* (Coccolithovirus, Schroeder et al., 2002) that outgroup all Phycodnaviruses sampled so far; and most importantly and (iii) even if these primers did not amplify existing Phycodnaviruses, it would hardly invalidate our findings regarding both, the phylodynamics of the detected demes and the presence of a barrier on the exchange of Phycodnaviruses among aquatic systems (that is, fresh and salt water), which may have more to do with host than virus biology. Nevertheless, a thorough study of Phycodnavirus phylodynamics in estuaries is essential to address hypotheses about salinity tolerances, and gene flow or transitions between freshwater and marine habitats. In conclusion, our preliminary coarse-grained phylodynamics and phylogeography analyses showed an interesting pattern of long-term dynamic fluctuation of viral demes, and strongly support the idea that there is no significant gene flow between terrestrial and oceanic aquatic environments, possibly because salinity restriction may be a determinant factor limiting Phycodnavirus survival.

Acknowledgements

We thank Alex Krusche (CENA, USP) for the logistical support, Telma A Monezi (ICB-USP) for technical support, Jonismar Souza, Alice Plaskievicz and Anthony Lopes (INPA), and Patricia Garrafa and Bernardo Gimenes (ICB-USP) for the valuable help during the sample collection and processing. We also thank Jérôme Payet, Caroline Chénard, Matthias Fischer, Jessica Labonté and Amy Chan (EOS-UBC) for their priceless assistance with specific protocols and techniques involved in the project. This project was financed by FAPESP (Grant no. 05/55164-1; scholarship no. 04/11287-0), supported by CAPES and in part by a Discovery Grant from the Natural Engineering and Research Council of Canada to CAS. PMAZ holds a CNPq PQ scholarship. We thank two anonymous referees for their meticulous reading and relevant contributions to the manuscript.

References

- Benner R, Opsahl S, Chin-Leo G, Richey JE, Forsberg BR. (1995). Bacterial carbon metabolism in the Amazon river system. *Limnol Oceanogr* **40**: 1262–1270.
- Bratbak G, Egge JK, Heldal M. (1993). Viral mortality of the marine alga *Emiliania huxleyi* (Haptophyceae) and termination of algal blooms. *Mar Ecol Prog Ser* **93**: 39–48.
- Brussaard CPD. (2004). Viral control of phytoplankton populations—a review. *J Eukaryot Microbiol* **51**: 125–138.
- Campos PF, Willerslev E, Sher A, Orlando L, Axelsson E, Tikhonov A et al. (2010). Ancient DNA analyses exclude humans as the driving force behind late Pleistocene musk ox (*Ovibos moschatus*) population dynamics. *PNAS* **107**: 5675–5680.
- Castberg T, Larsen A, Sandaa RA, Brussaard CPD, Egge JK, Heldal M et al. (2001). Microbial population dynamics and diversity during a bloom of the marine coccilithophorid *Emiliania huxleyi* (Haptophyta). *Mar Ecol Prog Ser* **221**: 39–46.
- Chen F, Suttle CA. (1995). Nested PCR with 3 highly degenerate primers for amplification and identification of DNA from related organisms. *Biotechniques* **18**: 609–612.
- Chen F, Suttle CA. (1996). Evolutionary relationships among large double-stranded DNA viruses that infect microalgae and other organisms as inferred from DNA polymerase genes. *Virology* **219**: 170–178.
- Cissoko MM, Desnues A, Bouvy M, Sime-ngando TL, Verling E, Bettarel Y. (2008). Effects of freshwater and seawater mixing on viro and bacterioplankton in a tropical estuary. *Freshwater Biol* **53**: 1154–1162.
- Clasen JL, Suttle CA. (2009). Identification of freshwater *Phycodnaviridae* and their potential phytoplankton hosts, using DNA pol sequence fragments and a genetic-distance analysis. *Appl Environ Microbiol* **75**: 991–997.
- Cottrell MT, Suttle CA. (1991). Wide-spread occurrence and clonal variation in viruses which cause lysis of a cosmopolitan, eukaryotic marine phytoplankton, *Micromonas pusilla*. *Mar Ecol Prog Ser* **78**: 1–9.
- Denman KL, Brasseur G, Chidthaisong A, Ciais P, Cox PM, Dickinson RE et al. (2007). Couplings between changes in the climate system and biogeochemistry. In: Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL (eds). *Climate Change 2007—The Physical Science Basis*. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press: Cambridge, UK; New York, NY, pp 499–587.
- Drake JW, Hwang CBC. (2005). On the mutation rate of Herpes Simplex Virus Type 1. *Genetics* **170**: 969–970.
- Drummond AJ, Rambaut A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* **7**: 214.
- Duffy S, Shackelton LA, Holmes EC. (2008). Rates of evolutionary change in viruses: patterns and determinants. *Nature Rev Genet* **9**: 267–276.
- Dunigan DD, Fitzgerald LA, Van Etten JL. (2006). Phycodnaviruses: a peek at genetic diversity. *Virus Res* **117**: 119–132.
- Felsenstein J. (1993). *PHYLIP (Phylogeny Inference Package) version 3.5c*. Distributed by the author. Department of Genetics, University of Washington: Seattle.

- Felsenstein J (ed.) (2004). *Inferring Phylogenies*. Sinauer: Sunderland.
- Ferguson N, Anderson R, Gupta S. (1999). The effect of antibody-dependent enhancement on the transmission dynamics and persistence of multiple-strain pathogens. *Proc Natl Acad Sci* **96**: 790–794.
- Gago S, Elena SF, Flores R, Sanjuán R. (2009). Extremely high mutation rate of a hammerhead viroid. *Science* **323**: 1308.
- Gobler CJ, Hutchins DA, Fisher NS, Cosper EM, Sanudo-Wilhelmy AS. (1997). Release and bioavailability of C, N, P, Se and Fe following viral lysis of a marine chrysophyte. *Limnol Oceanogr* **42**: 1492–1504.
- Gradstein FM, Ogg JG, Smith AG (eds) (2005). *A Geologic Time Scale*. Cambridge University Press: Cambridge.
- Grenfell BT, Pybus OG, Gog JR, Wood JLN, Daly JM, Mumford JA et al. (2004). Unifying the epidemiological and evolutionary dynamics of pathogens. *Science* **303**: 327–332.
- Guindon S, Gascuel O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**: 696–704.
- Hewson I, Capone DG, Steele JA, Fuhrmann JA. (2006). Influence of Amazon and Orinoco offshore surface water plumes on oligotrophic bacterioplankton diversity in the west tropical Atlantic. *Aquat Microb Ecol* **43**: 11–22.
- Hill RW, White BA, Cottrell MT, Dacey JWH. (1998). Virus-mediated total release of dimethylsulfoniopropionate from marine phytoplankton: a potential climate process. *Aquat Microb Ecol* **14**: 1–6.
- Houghton RA. (2007). Balancing the global carbon budget. *Annu Rev Earth Pl Sc* **35**: 313–347.
- Hungria M, Astolfi-Filho S, Chueire LMO, Nicolás MF, Santos EBP, Bulbul MR et al. (2005). Genetic characterization of *Chromobacterium* isolates from black water environments in the Brazilian Amazon. *Lett Appl Microb* **41**: 17–23.
- Jacquet S, Heldal M, Iglesias-Rodriguez D, Larsen A, Wilson W, Bratbak G. (2002). Flow cytometric analysis of an *Emiliania huxleyi* bloom terminated by viral infection. *Aquat Microb Ecol* **27**: 111–124.
- Kingman JFC. (1982). The coalescent. *Stoch Proc Appl* **13**: 235–248.
- Kirst CO. (1989). Salinity tolerance of eukaryotic marine algae. *Annu Rev Plant Physiol Plant Mol Biol* **40**: 21–53.
- Larsen JB, Larsen A, Bratbak G, Sandaa RA. (2008). Phylogenetic analysis of members of the *Phycodnaviridae* virus family, using amplified fragments of the major capsid protein gene. *Appl Environ Microbiol* **74**: 3048–3057.
- Lionard M, Muylaert K, Gansbeke DV, Vyverman W. (2005). Influence of changes in salinity and light intensity on growth of phytoplankton communities from the Schelde river and estuary (Belgium/The Netherlands). *Hydrobiologia* **540**: 105–115.
- Liss PS, Hatton AD, Malin G, Nightingale PD, Turner SM. (1997). Marine sulphur emissions. *Phil Trans R Soc Lond* **352**: 159–169.
- Logares R, Brate J, Bertilsson S, Clasen JL, Shalchian-Tabrizi K, Rengefors K. (2009). Infrequent marine-freshwater transitions in the microbial world. *Trends Microbiol* **17**: 414–422.
- Malin G, Wilson WH, Bratbak G, Liss PS, Mann NH. (1998). Elevated production of dimethylsulfide resulting from viral infection of cultures of *Phaeocystis pouchetii*. *Limnol Oceanogr* **43**: 1389–1393.
- Mehnert DU, Stewien KE. (1993). Detection and distribution of rotaviruses in raw sewage and creeks in São Paulo, Brazil. *Appl Environ Microbiol* **59**: 140–143.
- Page RDM, Holmes EC. (1998). *Molecular Evolution: A Phylogenetic Approach*. Blackwell Publishing Ltd: Oxford, UK, Chapter 4.5, pp 124–134.
- Pagel M, Meade A. (2004). A phylogenetic mixture model for detecting pattern-heterogeneity in gene sequence or character-state data. *Syst Biol* **53**: 571–581.
- Pybus OG, Rambaut A. (2009). Evolutionary analysis of the dynamics of viral infectious disease. *Nat Rev Genet* **10**: 540–550.
- Queiroz APS, Santos FM, Sassaroli A, Hársi CM, Monezi TA, Mehnert DU. (2001). Electropositive filter membrane as an alternative for the elimination of PCR inhibitors from sewage and water samples. *Appl Environ Microbiol* **67**: 4614–4618.
- Rejas D, Muylaert K, De Meester L. (2005). Nutrient limitation of bacteria and sources of nutrients supporting nutrient-limited bacterial growth in an Amazonian floodplain lake. *Aquat Microb Ecol* **39**: 57–67.
- Rio D, Sprovieri R, Castradori D, Di Stefano E. (1998). The Gelasian Stage (Upper Pliocene): a new unit of the global standard chronostratigraphic scale. *Episodes* **21**: 82–87.
- Romano CM, de Melo FL, Corsini MAB, Holmes EC, Zanotto PMA. (2007). Demographic histories of ERV-K in humans, chimpanzees and Rhesus monkeys. *PLoS One* **10**: e1026.
- Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, Yooseph S et al. (2007). The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biol* **5**: 398–431.
- Sabine CL, Feely RA, Gruber N, Key RM, Lee K, Bullister JL et al. (2004). The ocean sink for anthropogenic CO₂. *Science* **305**: 367–371.
- Sano E, Carlson S, Wegley L, Rohwer F. (2004). Movement of viruses between biomes. *Appl Environ Microbiol* **70**: 5842–5846.
- Schroeder DC, Oke J, Malin G, Wilson WH. (2002). Coccolithovirus (*Phycodnaviridae*): characterization of a new large dsDNA algal virus that infects *Emiliania huxleyi*. *Arch Virol* **147**: 1685–1698.
- Shapiro B, Rambaut A, Drummond AJ. (2006). Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. *Mol Biol Evol* **23**: 7–9.
- Short SM, Suttle CA. (2002). Sequence analysis of marine virus communities reveals that groups of related algal viruses are widely distributed in nature. *Appl Environ Microbiol* **68**: 1290–1296.
- Short SM, Suttle CA. (2003). Temporal dynamics of natural communities of marine algal viruses and eukaryotes. *Aquat Microb Ecol* **32**: 107–119.
- Short SM, Short CM. (2008). Diversity of algal viruses in various North American freshwater environments. *Aquat Microb Ecol* **51**: 13–21.
- Sioli H. (1975a). Tropical river as expressions of their terrestrial environments. In: Golley FB, Medina E (eds) *Tropical Ecological Systems Trends in Terrestrial and Aquatic Research*. Springer-Verlag: Berlin, pp 289–312.
- Sioli H. (1975b). Tropical river: the Amazon. In: Whitton, BA (ed) *River Ecology*. U.C. Press: Berkeley, pp 461–487.
- Suttle CA, Chan AM, Cottrell MT. (1991). Use of ultrafiltration to isolate viruses from seawater which

- are pathogens of marine phytoplankton. *Appl Environ Microbiol* **57**: 721–726.
- Thingstad TF. (2000). Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. *Limnol Oceanogr* **45**: 1320–1328.
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**: 4876–4882.
- Van Etten JL, Ghabrial SA. (1991). *Phycodnaviridae*. In: Franck RIB, Fauguet CM, Knudson DL, Brown F (eds). Classification and Nomenclature of Viruses, fifth report. Springer-Verlag: Vienna, pp 137–139.
- Van Etten JL, Graves MV, Müller DG, Boland W, Delaroche N. (2002). *Phycodnaviridae*—large DNA viruses. Springer-Verlag: Vienna, pp. *Arch Virol* **147**: 1479–1516.
- Wilhelm SW, Suttle CA. (1999). Viruses and nutrient cycles in the sea. *Bioscience* **49**: 781–788.
- Wilson WH, Schroeder DC, Ho J, Cantz M. (2006). Phylogenetic analysis of PgV-102P, a new virus from the English Channel that infects *Phaeocystis globosa*. *J Mar Biol Ass UK* **86**: 485–490.
- Wilson WH, Van Etten JL, Allen MJ. (2008). The Phycodnaviridae: the story of how tiny giants rule the world. In: Van Etten J (ed.) *Lesser Known Large dsDNA Viruses*. Springer: Berlin, pp 1–42.
- Zanotto PM, Gould EA, Gao GF, Harvey PH, Holmes EC. (1996). Population dynamics revealed by flaviviruses by molecular phylogenies. *Proc Natl Acad Sci* **93**: 548–553.
- Zhang YP, Suttle CA. (1994). Design and use of PCR primers for B-family DNA polymerase genes to detect and identify viruses and microbes. In: *ASLO & PSA Joint Meeting*. American Society for Limnology and Oceanography: Lawrence, Kans, p a–85.



This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0/>

Supplementary Information accompanies the paper on The ISME Journal website (<http://www.nature.com/ismej>)